SELECTION FOR AN OPTIMUM GROWTH CURVE

G. M. Tallis

The Johns Hopkins University

BU-223-M July, 1966

Abstract

In this paper, we consider the growth curves of animals. These may be regarded as the realization of a stochastic process $P(t)$ with expectation function $\mu(t)$, $t \geq 0$. Suppose now there is an optimum curve $\alpha(t)$, then the problem is to construct a selection index, I, which will gradually move $\mu(t)$ to $\alpha(t)$ at some or all points $t \in [0,a]$.

Biometrics Unit, Plant Breeding Department, Cornell University, Ithaca, New York.

SELECTION FOR AN OPTIMUM GROWTH CURVE

G. M. Tallis
The Johns Hopkins University

BU-223-M July, 1966

This work was supported in part by Research Grant GM-13225 from the National Institutes of Health while the author was a Visiting Associate Professor in the Biometrics Unit, Plant Breeding Department, Cornell University, in July-August, 1966.

SELECTION FOR AN OPTIMUM GROWTH CURVE

G. M. Tallis
The Johns Hopkins University

In this paper, we consider the growth curves of animals. These may be regarded as the realization of a stochastic process $P(t)$ with expectation function $\mu(t)$, $t \geq 0$. Suppose now there is an optimum curve $\alpha(t)$, then the problem is to construct a selection index, I, which will gradually move $\mu(t)$ to $\alpha(t)$ at some or all points $t \in [0,a]$.

Discrete Solution

We firstly consider the problem of changing $\mu(t)$ to $\alpha(t)$ at well defined points t_i, $i=1,2,\cdots,n$. These points may be considered as defining critical stages in the growth cycle of the animals. In order to make progress, we assume the usual additive model $P(t) = \mu(t) + g(t) + e(t)$ where $g(t)$ is the additive genetic and $e(t)$ the environmental contributions to phenotype at time t. The components $g(t)$ and $e(t)$ are considered to satisfy the relation $E\{g(s)\, e(t)\} = 0$ for all $s,t \in [0,a]$.

For convenience we write $p_i \equiv p(t_i) \equiv P(t_i) - \mu(t_i)$, $\mu(t_i) \equiv \mu_i$, $g_i \equiv g(t_i)$ and $e_i \equiv e(t_i)$. In this notation then the problem is to construct a selection index $I = \sum_{i=1}^{n} \beta_i p_i = \underset{\sim}{\beta}' \underset{\sim}{p}$ such that $\mathrm{Cov}(g_i, I) = k_i$, $i=1,2,\cdots,n$ where $k_i = (\alpha_i - \mu_i)$. Since under standard assumptions the expected change in μ_i, Δ_i, after selection on I is $\Delta I\, \mathrm{Cov}(g_i, I)/\sigma_I^2$, where Δ_I is the selection differential in I, suitably defined, we would have that $\Delta_i = \Delta_I k_i/\sigma_I^2$. Hence, after approximately σ_I^2/Δ_I selections the total change in μ_i should be about k_i and the optimum

values will have been reached.

In order to find $\underset{\sim}{\beta}$ notice that the above conditions can be written in matrix form as $E\{\underset{\sim}{g}\underset{\sim}{p}'\underset{\sim}{\beta}\} = \underset{\sim}{k}$ and, since $E\{\underset{\sim}{g}\underset{\sim}{p}'\} = \underset{\sim}{G}$, the genetic covariance matrix, $\underset{\sim}{\beta} = \underset{\sim}{G}^{-1}\underset{\sim}{k}$. This is a special case of a restricted selection index reported by the author in (1962).

For a rather wide class of procedures for estimating $\underset{\sim}{G}$, $\hat{\underset{\sim}{\beta}} = \hat{\underset{\sim}{G}}^{-1}\underset{\sim}{k}$ is a consistent estimator of $\underset{\sim}{\beta}$. Moreover, the analysis can be put in the multivariate analysis of covariance form

Source	d.f.	Estimate	Expectation
Between families	f_b	$\hat{\underset{\sim}{B}}$	$\underset{\sim}{B} = \underset{\sim}{W} + (r/m)\underset{\sim}{G}$
Error	f_w	$\hat{\underset{\sim}{W}}$	$\underset{\sim}{W}$

where r is the number per family group and m=2 or m=4 according to the type of family relationship. Thus, $\hat{\underset{\sim}{G}} = \frac{m}{r}[\hat{\underset{\sim}{B}}-\hat{\underset{\sim}{W}}]$.

The covariance matrix for $\hat{\underset{\sim}{\beta}}$, $\underset{\sim}{V}(\hat{\underset{\sim}{\beta}})$, is obtained by noticing that $\underset{\sim}{G}(d\underset{\sim}{\beta})$ $+ (d\underset{\sim}{G})\underset{\sim}{\beta} = \underset{\sim}{0}$, whence $\underset{\sim}{G}(d\underset{\sim}{\beta}) = -(d\underset{\sim}{G})\underset{\sim}{\beta}$. Thus, proceeding in the standard fashion

$$\underset{\sim}{G}\underset{\sim}{V}(\hat{\underset{\sim}{\beta}})\underset{\sim}{G} = E\{d\underset{\sim}{G}\underset{\sim}{\beta}\underset{\sim}{\beta}'d\underset{\sim}{G}\} \ , \qquad (1)$$

and letting $\underset{\sim}{\beta}\underset{\sim}{\beta}' = \underset{\sim}{\gamma}$, then the q,rth element of $\underset{\sim}{G}\underset{\sim}{V}(\hat{\underset{\sim}{\beta}})\underset{\sim}{G}$ is $\sum_s \sum_t \gamma_{st}\ \mathrm{Cov}(\hat{G}_{tr},\hat{G}_{qs})$. But, it can be shown that

$$\mathrm{Cov}(\hat{G}_{tr},\hat{G}_{qs}) = \frac{m^2}{r^2}\left\{(B_{tq}B_{rs}+B_{ts}B_{rq})f_b^{-1} + (W_{tq}W_{rs}+W_{ts}W_{rq})f_w^{-1}\right\}$$

and substituting this into (1) and simplifying

$$\underset{\sim}{V}(\hat{\underset{\sim}{\beta}}) \simeq \frac{m^2}{r^2}\, \underset{\sim}{G}^{-1}\left\{(\underset{\sim}{B}\underset{\sim}{\gamma}\underset{\sim}{B}+\underset{\sim}{B}\ \text{Trace}\ \underset{\sim}{B}\underset{\sim}{\gamma})f_b^{-1} + (\underset{\sim}{W}\underset{\sim}{\gamma}\underset{\sim}{W}+\underset{\sim}{W}\ \text{Trace}\ \underset{\sim}{W}\underset{\sim}{\gamma})f_w^{-1}\right\}\underset{\sim}{G}^{-1}. \qquad (2)$$

It is now possible to calculate the approximate variance of the estimate $\hat{I} = \hat{\underset{\sim}{\beta}}'\underset{\sim}{p}$. We write $\hat{\underset{\sim}{\beta}} = \underset{\sim}{\beta} + \Delta\underset{\sim}{\beta}$, then since $E\{\Delta\underset{\sim}{\beta}\underset{\sim}{p}'\} = \underset{\sim}{0}$, $E(\hat{I}) = 0$ and

$$V(\hat{I}) = E(\underset{\sim}{\beta}'\underset{\sim}{p}+\Delta\underset{\sim}{\beta}'\underset{\sim}{p})^2 = \underset{\sim}{\beta}'\underset{\sim}{P}\underset{\sim}{\beta} + \text{Trace}(\underset{\sim}{V}(\hat{\underset{\sim}{\beta}})\underset{\sim}{P}) \qquad (3)$$

where $\underset{\sim}{P} = E(\underset{\sim}{p}\underset{\sim}{p}')$, the phenotypic covariance matrix.

The Continuous Solution

Although the index constructed above, $\underset{\sim}{\beta} = \underset{\sim}{G}^{-1}\underset{\sim}{k}$, may be of some practical importance, it is only of minor theoretical interest. What is really required is some index which will apply selection pressure to all points of the curve simultaneously. In this section we propose to discuss such an index.

We introduce now the covariance function $E\{p(s)p(t)\} = \Gamma_p(s,t) = \Gamma_g(s,t) + \Gamma_e(s,t)$, and in the following argument $\Gamma_g(s,t)$ will replace the matrix $\underset{\sim}{G}$ of the discrete solution. For this treatment we let $k(t) = [\alpha(t)-\mu(t)]$, a continuous function of time. Our index, instead of being of the form $\underset{\sim}{\beta}'\underset{\sim}{p}$, is $I = \int_0^a \beta(t)p(t)dt$, where the integral is to be interpreted as a stochastic integral. It is well known that a sufficient condition for the latter to exist is that $\int_0^a\int_0^a \Gamma_p(s,t)\beta(s)\beta(t)\ dsdt = \sigma_I^2$ exists.

The condition that $Cov(g_i,I) = k_i$ is now replaced by $E(g(t)I) = k(t)$. But, $E\{g(t)\int_0^a p(s)\beta(s)ds\} = \int_0^a \beta(s)\Gamma_g(s,t)ds = k(t)$ (see appendix), and in order to find $\beta(s)$ the integral equation

$$\int_0^a \beta(s)\Gamma_g(s,t)ds = k(t) \qquad (4)$$

must be solved. Equation (4) is a Fredholm integral equation of the first kind and inversion, in general, is not an easy task.

If $p(t)$ is a normal process, then $\int_0^a p(t)\beta(t)dt = I$ will also be normal (Loève, 1955, page 485). Thus the expected change at the point $t \in [0,a]$ after selection will be $\Delta(t) = \Delta_I k(t)/\sigma_I^2$. Hence, as for the discrete case, after σ_I^2/Δ_I selections, $\mu(t)$ should be near $\alpha(t)$ at all points $t \in [0,a]$.

Some points pertaining to the solution of the selection integral equation will now be discussed. Firstly, in order to make some progress, some parametric form must be assigned to the covariance kernel $\Gamma_g(s,t)$ and, for purposes of illustration, we will let $\Gamma_g(s,t) = \sum_{i=1}^n \omega_i e^{\eta_i st}$, ω_i and η_i real, $\omega_i > 0$. Although the form of Γ_g has been chosen for mathematical convenience, this model should be satisfactory whenever the kernel can be assumed strictly positive.

In order to find a solution to (4) one can sttempt to write $\beta(s) = \sum_{i=0}^{\infty} \theta_i\phi_i(s)$, where the θ_i are real constants and the $\phi_i(s)$ are the eigenfunctions of the kernel $\Gamma_g(s,t)$. Howevêr, with the particular form assumed above we can calculate an approximation to $\beta(s)$ in another way.

We have $\int_0^a \left(\sum_{i=1}^n \omega_i e^{\eta_i st}\right)\beta(s)ds = k(t)$ and, by differentiating both sides

with respect to t and setting t=0, we obtain $\mu_1 \left(\sum_{i=1}^{n} \omega_i \eta_i \right) = k'(0)$ where $\mu_1 = \int_0^a s\beta(s)ds$. Similarly $\mu_j \left(\sum_{i=1}^{n} \omega_i \eta_i^j \right) = k^{(j)}(0)$, $\mu_j = \int_0^a s^j \beta(s)ds$, and provided the $k^{(j)}(0)$ exist these equations can be solved for μ_j if $\sum_{i=1}^{n} \omega_i \eta_i^j \neq 0$.

If, now, $\beta(s)$ can be expressed approximately as a polynomial of degree m-1, $\beta(s) = \sum_{i=0}^{m-1} b_i s^i$, then $\mu_j = \sum_{i=0}^{m-1} b_i a^{i+j+1}/(j+i+1)$ and the vector of coefficients, $\underset{\sim}{b}$, can be found from the equation

$$\underset{\sim}{b} = \underset{\sim}{A}^{-1} \underset{\sim}{\mu}$$

where $\underset{\sim}{A} = [a^{i+j+1}/(i+j+1)]$ and $\underset{\sim}{\mu}' = (\mu_1, \mu_2, \cdots, \mu_m)$. A higher degree polynomial $\beta(s)$ can be constructed by taking more moments and, if for some j $\Sigma \omega_i \eta_i^j = 0$, then the j^{th} moment is omitted and the process is completed using an additional moment of higher order than μ_j.

From the Weierstrass approximation theorem it follows that if a continuous solution to (4) exists, then the above procedure leads to a uniformly close approximation to $\beta(s)$. That is, given an ϵ there exists an m=M such that for all $m > M$, $\left| \sum_{i=1}^{m-1} b_i s^i - \beta(s) \right| < \epsilon$ for all $s \in [0,a]$.

In order to obtain estimates of the parameters ω_i and η_i, $\hat{\omega}_i$ and $\hat{\eta}_i$, i=1,2,$\cdots$,n, we may obtain estimates of genetic covariance, $\mathrm{Cov}(g(t_q), g(t_r)) = G_{qr}$, as described earlier, and use these in a standard non-linear least squares analysis to calculate $\hat{\omega}_i$ and $\hat{\eta}_i$. The details of such an analysis will not be given here since they are well established.

In the above case it is not difficult to obtain an expression for $V(\hat{I})$ which is entirely analagous to (3). It will be possible, in general, to obtain approximate sampling variances and covariances for the $\hat{\omega}_i$ and $\hat{\eta}_i$; hence we can find $\underset{\sim}{V}(\hat{\underset{\sim}{\mu}})$. Now, $\hat{\underset{\sim}{b}} = \underset{\sim}{A}^{-1}\hat{\underset{\sim}{\mu}}$, $\hat{\beta}(s) = \hat{\underset{\sim}{b}}'\underset{\sim}{s}$ when $\underset{\sim}{s}' = [1, s, s^2, \cdots, s^{m-1}]$ and $V(\hat{\underset{\sim}{b}}) = \underset{\sim}{A}^{-1}\underset{\sim}{V}(\hat{\underset{\sim}{\mu}})\underset{\sim}{A}^{-1}$.

If we let $\hat{\beta}(s) = \beta(s) + \Delta\beta(s)$, then

$$V(\hat{I}) = V\left(\int_0^a p(t)\hat{\beta}(t)dt\right)$$

$$= E\left\{\int_0^a p(t)\left(\beta(t)+\Delta\beta(t)\right)dt \cdot \int_0^a p(s)\left(\beta(s)+\Delta\beta(s)\right)ds\right\}$$

and using the assumption that $E\{p(t)\Delta\beta(s)\} \equiv 0$ we find that

$$V(\hat{I}) = \int_0^a\int_0^a \beta(s)\beta(t)\Gamma_p(s,t)dsdt + \int_0^a\int_0^a B(s,t)\Gamma_p(t,s)dsdt \qquad (5)$$

where $B(s,t) = E\{\Delta\beta(s)\Delta\beta(t)\}$.

It is clear that in the example $B(s,t) = \underset{\sim}{s}'\underset{\sim}{A}^{-1}\underset{\sim}{V}(\hat{\underset{\sim}{\mu}})\underset{\sim}{A}^{-1}\underset{\sim}{t}$ and hence we have the required extension of (3) to the continuous case.

Extensions

We consider next the continuous analogue of the general genetic selection index. In the discrete case the problem is to find a vector $\underset{\sim}{\beta}$ such that $\underset{\sim}{\beta}'\underset{\sim}{p}$ is in some sense the best predictor of $\underset{\sim}{a}'\underset{\sim}{g}$, when $\underset{\sim}{a}$ is a vector of economic weights. By using a least squares argument, we minimize

$$E\{(\underset{\sim}{g}'\underset{\sim}{a}-\underset{\sim}{p}'\underset{\sim}{\beta})'(\underset{\sim}{g}'\underset{\sim}{a}-\underset{\sim}{p}'\beta)\}$$

$$= \underset{\sim}{a}'\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a} - 2\underset{\sim}{\beta}'\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a} + \underset{\sim}{\beta}'\underset{\sim}{\underset{\sim}{P}}\underset{\sim}{\beta} \tag{6}$$

$$= \underset{\sim}{a}'\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a} - \underset{\sim}{a}'\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{\underset{\sim}{P}}^{-1}\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a} + (\underset{\sim}{\beta}-\underset{\sim}{\underset{\sim}{P}}^{-1}\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a})'\underset{\sim}{\underset{\sim}{P}}(\underset{\sim}{\beta}-\underset{\sim}{\underset{\sim}{P}}^{-1}\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a})$$

with respect to $\underset{\sim}{\beta}$. It is clear that equation (6) is at a minimum when $\underset{\sim}{\beta} = \underset{\sim}{\underset{\sim}{P}}^{-1}\underset{\sim}{\underset{\sim}{G}}\underset{\sim}{a}$, which is the required solution.

Introducing the economic weight function a(t), in the continuous case we must minimize

$$I(\beta) = E\left\{\int_0^a g(t)a(t)dt - \int_0^a p(t)\beta(t)dt\right\}^2$$

$$= \int_0^a\int_0^a a(s)a(t)\Gamma_g(s,t)dsdt$$

$$- 2\int_0^a\int_0^a a(t)\beta(s)\Gamma_g(s,t)dsdt$$

$$+ \int_0^a\int_0^a \beta(s)\beta(t)\Gamma_p(s,t)dsdt$$

with respect to $\beta(t)$. This is a variational problem and we write $I(\beta+\epsilon\xi) = F(\epsilon)$, where $\xi(t)$ is an arbitrary continuous function vanishing at 0 and a. Then

$$\delta I = 0 = \epsilon\,\frac{dF(0)}{d\epsilon} = 2\epsilon\int_0^a \xi(t)\left[\int_0^a \Gamma_g(s,t)a(s)ds - \int_0^a \Gamma_p(s,t)\beta(s)ds\right]dt$$

implies, since $\xi(t)$ is arbitrary, that

$$\int_0^a \Gamma_p(s,t)\beta(s)ds = \int_0^a \Gamma_g(s,t)a(s)ds \quad . \tag{7}$$

Thus, (7) is another Fredholm integral equation of the first kind which must be solved for $\beta(t)$.

Finally, suppose that, instead of just a single growth curve, we are interested in different curves which describe the overall phenotypic growth of the animal. Thus, we have a vector valued random process $\underset{\sim}{P}(t)' = [P_1(t),P_2(t),\cdots,P_q(t)]$ and we let

$$E\{\underset{\sim}{P}(t)\} = \underset{\sim}{\mu}(t) \ , \ E\{[P(t)-\mu(t)][P(s)-\mu(s)]'\} = \Gamma_p(s,t) = [\Gamma_p^{ij}(s,t]$$

where

$$\Gamma_p^{ij}(s,t) = E\{p_i(t)p_j(s)\} = \Gamma_g^{ij}(s,t) + \Gamma_e^{ij}(s,t) \ .$$

The problem now is to move $\underset{\sim}{\mu}(t)$ to $\underset{\sim}{\mu}(t) + \underset{\sim}{k}(t)$.

Consider $I = \sum_{j=1}^{q} \int_0^a p_j(s)\beta_j(s)ds$ as the prosptective index. Then we must have

$$E\{g_i(t)I\} = k_i(t) = \sum_{j=1}^{q} \int_0^a \Gamma_g^{ij}(s,t)\beta_j(s)ds \ , \quad i=1,2,\cdots,$$

or in obvious matrix form

$$\int_0^a \underset{\sim}{\Gamma}_g(s,t)\underset{\sim}{\beta}(s)ds = \underset{\sim}{k}(t) \quad . \tag{8}$$

In the finite case these results specialize as follows. Let $I = \sum_{j=1}^{q} \underset{\sim}{p}_j'\underset{\sim}{\beta}_j$, where $\underset{\sim}{p}_j' = [p_j(t_1),p_j(t_2),\cdots,p_j(t_n)]$, then the condition is that

$$E\{\underset{\sim}{g}_i I\} = \underset{\sim}{k}_i \ , \ k_i' = [k_i(t_1),k_i(t_2),\cdots,k_i(t_n)] \quad .$$

If $G(ij) = E\{g_i g_j'\}$, then we obtain the equation

$$\sum_{j=1}^{q} \underset{\sim}{G}(ij)\underset{\sim}{\beta}_j = \underset{\sim}{k}_i \quad ; \qquad i=1,2,\cdots,q$$

or, in full matrix notation

$$\underset{\sim\sim}{G}\underset{\sim}{\beta} = \underset{\sim}{k} \tag{9}$$

where $\underset{\sim}{\beta}' = [\underset{\sim}{\beta}_1', \underset{\sim}{\beta}_2', \cdots, \underset{\sim}{\beta}_q']$, $\underset{\sim}{k}' = [\underset{\sim}{k}_1', \underset{\sim}{k}_2', \cdots, \underset{\sim}{k}_q']$ and

$$G = \begin{bmatrix} \underset{\sim}{G}_{11} & \underset{\sim}{G}_{12} & \cdots & \underset{\sim}{G}_{1q} \\ \underset{\sim}{G}_{21} & \underset{\sim}{G}_{22} & \cdots & \underset{\sim}{G}_{2q} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{G}_{q1} & \underset{\sim}{G}_{q2} & \cdots & \underset{\sim}{G}_{qq} \end{bmatrix} .$$

From (9) it is clear that a unique solution to the more general problem is guaranteed.

Discussion

There are a number of practical situations to which the techniques of the previous sections could be applied. For instance, in fat lamb production one desirable characteristic is rapid and early increases in body weight. On the other hand, if these weight increments are too great then some production and marketing problems may arise. Hence, some optimal growth curve can perhaps

be specified towards which the average flock performance is to be pushed.

Different breeds of livestock are characterized not only by the quality of the associated primary product, but also by how rapidly, in what quantity and how efficiently it is produced. Thus, early maturity in lambs may be a desirable feature under some systems of management and marketing, whereas under entirely different conditions late maturing sheep may be optimal. Under any specific set of circumstances it makes sense to use the particular breed which has, among other things, the correct growth pattern. The concept of an optimal growth curve, therefore, appears to have genuine and important practical implications.

In most cases one feels that the discrete solution suggested above would provide an adequate selection tool. Practically, one would probably be satisfied to have the population growth curve approach the optimal curve at a finite number of points since intuition suggests that intermediate points will also be brought near optimality automatically. Provided n is a reasonable number, the vector $\underset{\sim}{\beta}$ can be estimated with little trouble.

However, it is not just of academic interest to investigate possible means of applying selection pressure throughout a continuum. It is worthwhile to see where the theory extends and to observe the similarity between the general and the specific results. If very many measurements were recorded on each individual it is possible that the discrete methods would become unwieldy, and the continuous solution may be the most appropriate approximation to use. Of course, with enough points numerical quadrature methods could be used to accurately estimate $\hat{I} = \int_0^a \hat{\beta}(t)p(t)dt$.

Whatever may be the theoreticad advantages or disadvantages of the methods outlined here, experimental verification of their practical utility is required. As usual, this may be a long time in coming, but now at least there are methods to test.

Acknowledgement

The author is indebted to Dr. Earle Klosterman of the Ohio Agricultural Experiment Station, Wooster, for a discussion leading to the formulation of the problems discussed above.

References

Loève, M. (1963). Probability Theory. D. VanNostrand Company, Inc., Princeton, New Jersey.

Tallis, G. M. (1962). A selection index for optimum genotype. Biometrics 18: 120-122.

Appendix

Prior to equation (3) it is asserted that $E\{g(t)I\} = \int_0^a \beta(s)\Gamma_g(s,t)ds$ and the purpose here is to establish this result. Let $t_0,t_1,t_2,\cdots,t_n$ be a partition of the interval $[0,a]$ and define

$$I_n = \sum_{i=1}^{n} \beta(t_i)p(t_i)(t_i-t_{i-1})$$

then

$$|E\{g(t)(I-I_n)\}| \le E\{|g(t)(I-I_n)|\} \le [E\{g^2(t)\}\cdot E\{(I-I_n)^2\}]^{\frac{1}{2}}$$

by the Schwartz inequality. By taking limits on both sides of the above inequality we find, since $\lim_{n\to\infty} E\{I-I_n)^2\} = 0$, that

$$E\{g(t)I\} = \lim_{n\to\infty} \sum_{i=1}^{n} \beta(s_i)\Gamma_g(s_i,t)(s_i-s_{i-1})$$

$$= \int_0^a \beta(s)\Gamma_g(s,t)ds \quad .$$

The last integral exists by assumption.